The Effects of Differentiated Instruction and Enrichment Pedagogy on Reading Achievement in Five Elementary Schools

Sally M. Reis, D. Betsy McCoach, Catherine A. Little, Lisa M. Muller and R. Burcu Kaniskan

Am Educ Res J 2011 48: 462 originally published online 4 October 2010
DOI: 10.3102/0002831210382891

American Educational Research Journal
April 2011, Vol. 48, No. 2, pp. 462–501
DOI: 10.3102/0002831210382891

The Effects of Differentiated Instruction and Enrichment Pedagogy on Reading Achievement in Five Elementary Schools

Sally M. Reis
D. Betsy McCoach
Catherine A. Little
Lisa M. Muller
R. Burcu Kaniskan
University of Connecticut

This experimental study examined the effect of a differentiated, enriched reading program on students' oral reading fluency and comprehension using the schoolwide enrichment model–reading (SEM-R). Treatment and control conditions were randomly assigned to 63 teachers and 1,192 second through fifth grade students across five elementary schools. Using multilevel modeling, significant differences favoring the SEM-R were found in reading fluency in two schools (Cohen's d *effect sizes of .33 and .10) and in reading comprehension in the high-poverty urban school (Cohen's* d *= .27), with no achievement differences in the remaining schools. These results demonstrate that an enrichment reading approach, with differentiated instruction and less whole group instruction, was as effective as or more effective than a traditional whole group basal approach.*

KEYWORDS: enrichment, engagement, differentiation, reading achievement

Reading achievement has long been acknowledged as a critical factor in school success (Anderson, Hiebert, Scott, & Wilkinson, 1985; Chall, Jacobs, & Baldwin, 1990; National Reading Panel [NRP], 2000). Results of current reading assessments continue to raise concerns, however, about reading achievement in the United States. In the 2007 National Assessment of Educational Progress, although both fourth and eighth grade results showed higher percentages of students performing at or above the basic level than in the past, only about a third of students were performing at or above the proficient level in reading (National Center for Education Statistics, 2007). Moreover, a significant gap continues between the reading performance of students from higher and lower socioeconomic levels.

In addition to general concerns about reading competency, a related concern among reading educators and researchers is students' declining interest and engagement in reading across all grade levels (Greenberg, Gilbert, & Fredrick, 2006; Guthrie & Wigfield, 2000; Pitcher et al., 2007; Unrau & Schlackman, 2006; Wigfield & Guthrie, 1997). One reason for this declining interest may be the frequent mismatch between the needs of students who read at different levels and the instructional opportunities provided to them, in terms of both reading instruction and time spent reading (Ivey & Broaddus, 2001; Pitcher et al., 2007; Reis et al., 2004). Although differentiated reading instruction is widely suggested in response to these learners' needs (Reis et al., 2004; Tomlinson, 1999, 2003), the implementation of differentiation continues to be limited at best, given the attention focused on remediation and supporting struggling learners (Latz, Speirs Neumeister, Adams, & Pierce, 2009; Moon, Brighton, & Callahan, 2003; Tomlinson, 2003). Even when advanced reading materials are available for above grade level readers, research has shown that students are rarely encouraged or guided to pursue them because of lack of teacher time and greater attention paid to the needs of readers who read well below grade level (Reis et al., 2004). In fact, despite the wide appeal and broad-base professional

---

SALLY M. REIS is Board of Trustees Distinguished Professor and Teaching Fellow in the Educational Psychology Department, Neag School of Education, University of Connecticut, 2131 Hillside Road, Unit 3007, Storrs, CT 06269-3007; e-mail: *sally.reis@uconn.edu*. Her research interests include giftedness and talent development, differentiation of curriculum and instruction, talented readers, and extensions of the schoolwide enrichment model.

D. BETSY MCCOACH is an associate professor of measurement, evaluation, and assessment in the Educational Psychology Department, Neag School of Education, University of Connecticut; e-mail: *betsy.mccoach@uconn.edu*. Her research interests include multilevel modeling, structural equation modeling, longitudinal data analysis, affective instrument design, and gifted education.

CATHERINE A. LITTLE is an associate professor in gifted and talented education, Educational Psychology Department, Neag School of Education, University of Connecticut; e-mail: *catherine.little@uconn.edu*. Her research interests include professional development, teacher talent development, and differentiation of curriculum and instruction.

LISA M. MULLER is a program specialist at the Neag Center for Gifted Education and Talent Development in the Educational Psychology Department at the University of Connecticut; e-mail: *lisa.muller@uconn.edu*. Her research interests include at-risk youth, differentiated reading instruction, and gifted education.

R. BURCU KANISKAN is a PhD student in the Educational Psychology Department, Neag School of Education, University of Connecticut; e-mail: *burcukaniskan@gmail.com*. Her research interests include latent variable modeling, multilevel modeling, longitudinal analysis, item response theory, value-added models, school effectiveness evaluation programs, differential item functioning, measurement invariance and parameter bias, and assessment of model fit.

development about differentiated instruction and some teachers' attempts to modify and differentiate language arts curriculum (Allington, 2002; Pressley, Wharton-McDonald, Mistretta, & Echevarria, 1998), research has shown that differentiation strategies are inconsistently implemented in many reading classrooms and are rarely used across the country (Archambault et al., 1993; Reis et al., 1993; Reis et al., 2004; Westberg, Archambault, Dobyns, & Salvin, 1993). In one experimental study about instructional and curriculum differentiation, the effects of using a differentiation strategy called curriculum compacting was examined to modify the curriculum and eliminate previously mastered work for high-ability students (Reis et al., 1993). The work that was eliminated was repeated content from previous textbooks or content that may be new in the curriculum but that some students had already mastered. When classroom teachers eliminated between 40% and 50% of the previously mastered regular curriculum for higher achieving students, no differences were found between students whose work was compacted and students who did *all* the work in reading, math computation, social studies, and spelling.

The premise of this study is similar, as it investigated the effects of an enrichment and differentiated approach to reading instruction, the schoolwide enrichment model–reading framework (SEM-R; Reis et al., 2009), on elementary students' reading fluency and comprehension in five elementary schools across the country. In particular, this study investigated whether the use of engagement and differentiation strategies and the elimination of up to 5 hours of whole group instruction each week produced higher, similar, or lower reading scores for students who participated in the intervention as compared to those who did not. The schools represent different geographic regions and serve students of varying backgrounds and achievement levels, including rural, urban, and suburban schools. One of the five was a low-scoring urban school (Urban Southeast) serving a high percentage of children who live in poverty. The SEM-R was implemented in these five schools from mid-September 2007 until the end of February 2008, for a total of 24 weeks. Teachers and students were randomly assigned to either the treatment or control group; treatment group classes completed an hour of SEM-R implementation daily, while control group classes continued with regular reading instruction for the same span of time.

The SEM-R is an enrichment-based reading program designed to stimulate interest in and enjoyment of reading, leading to higher reading achievement, by enabling students to self-select and read high-interest books of personal choice that are slightly to moderately above current reading instructional levels independently with differentiated instruction provided in weekly teacher conferences. Additional components of the model include (a) broad-based exposure to books through teacher read-alouds, (b) emphasis on self-regulation of reading and behavior, and (c) opportunities for students to engage in a wide variety of enrichment activities to extend their reading and pursue interests. All of these components are organized into

a three-phase structure, with phase lengths varying over time in response to student capacity to select books for independent reading time.

This research study was designed to examine the use of differentiated instruction and the effects of the use of the SEM-R on students' reading fluency and comprehension. In previous experimental studies, the SEM-R has been shown to be effective at increasing reading comprehension and fluency scores for elementary students in a smaller, geographically limited sample of schools, with results suggesting that urban, high-poverty students benefitted from this intervention (Reis et al., 2007; Reis et al., 2008; Reis & Housand, 2009). This study extended previous research by increasing the geographic and demographic diversity of schools, the number of students involved, and the length of the intervention to 24 weeks and by decreasing direct involvement from the research team in supporting classroom implementation of the intervention. Research team members continued to provide support for intervention, but at a decreased level of intensity compared to previous studies, in an effort to assess whether this differentiated, enrichment-based reading intervention could be implemented without the regular presence of research team members.

Theoretical Framework and Review of Research

Two major theoretical influences underlie the SEM-R framework. The first influence is the use of enrichment pedagogy as a way to increase engagement in reading, and the second is differentiation of content and instruction.

Theoretical Framework

The SEM-R applies a widely used enrichment program, the SEM, to reading. The three-phase structure of the SEM-R approach is derived from Renzulli's (1977) enrichment triad model (SEM), with three levels of enrichment: Type I (exposure), Type II (differentiated training in specific thinking and process skills), and Type III (investigations of self-selected topics). The SEM-R's three phases follow this learning approach, as Phase 1 focuses on exposing students to books, Phase 2 incorporates differentiated instruction, including specific reading strategy instruction, applied to self-selected independent reading, and Phase 3 allows students to pursue self-selected enrichment activities and projects related to reading.

The triad model, along with its larger-scale translation into the SEM (Renzulli, 1977; Renzulli & Reis, 1985; 1997), is one of the most popular approaches in gifted education pedagogy (VanTassel-Baska & Brown, 2007) and has also been used as a magnet and charter school theme with students in urban, suburban, and rural schools across the country for the past three decades (Reis & Renzulli, 2003; Renzulli & Reis, 1994). The SEM is widely used as an enrichment theme in both gifted and regular education

programs, with this broad applicability of the SEM's three central goals: developing talents in all children, providing a broad range of differentiated learning experiences for all students, and providing follow-up advanced learning opportunities for children based on abilities and interests. The SEM emphasizes the use of engaging and challenging learning experiences constructed around students' interests, learning styles, and product styles. Separate studies on the SEM have demonstrated its effectiveness in schools with widely differing socioeconomic levels and program organization patterns (Olenchak, 1988; Olenchak & Renzulli, 1989). The effectiveness of the model has been studied in more than 30 years of research and field testing, most recently in the use of SEM as a curricular framework (Reis & Fogarty, 2006; Reis, Gentry, & Maxfield, 1998; Reis et al., 2005).

A major goal of the SEM is engagement of students in self-selected learning opportunities, representing another component that has been integrated into the SEM-R. Current research connects increased levels of student engagement to increased achievement in reading (Guthrie & Wigfield, 2000; Teale & Gambrell, 2007), increased student motivation for reading (Gambrell, Palmer, Codling, & Mazzoni, 1996), achievement goals (Meece & Miller, 1999), and interest (Guthrie, Hoa, Wigfield, Tonks, & Perencevich, 2006). Guthrie's (2004) research suggests that engagement and enjoyment in reading may emerge when readers spend time reading and employ strategic cognitive behaviors that enable them to create meaning from text. A relationship has also been suggested between engagement and motivation as students who read more generally have higher motivation to read (Guthrie, 2004; Guthrie et al., 2006a; Guthrie et al., 2007) and may also have higher reading achievement (Reis et al., 2007; Taylor, Frye, & Maruyama, 1990). Recommended instructional practices to increase reading motivation and comprehension from Guthrie and Wigfield's research are embedded in the SEM-R, such as supporting student autonomy (Phase 2 and 3), exposure to and having students read interesting texts (Phase 1 and 2), facilitating social interactions related to reading (all phases), and maintaining strong relations between teachers and students (all phases; Guthrie et al., 2006b).

A second theoretical influence on the SEM-R is differentiated instruction using assessment data to support modification of curriculum and instruction to respond to differences in students' readiness, interests, and learning profiles (Renzulli, 1988; Tomlinson, 2001). Differentiated instruction emphasizes that learning is most effective when teachers are able to assess students' current levels of functioning and learning preferences and then use this information to help students progress to more advanced levels of functioning and a better match of learning opportunities. Differentiated instruction can be used to ensure that all students receive appropriate academic challenge as well as to promote engagement and higher achievement (e.g., Byrnes, 1996; Renzulli, 1977). Although

differentiated instruction is widely discussed as a goal in schools across the country and continues to be a national focus in professional development efforts, little experimental research has been conducted on its use, and teachers appear to struggle to implement differentiated instruction, facing challenges such as concerns about planning for and managing differentiation as well as fear of state assessments and little administrative support (Hertberg-Davis & Brighton, 2006; Latz et al., 2009; Moon et al., 2003; Reis et al., 1993; Reis et al., 2004; VanTassel-Baska & Stambaugh, 2005).

Research Related to the SEM-R Intervention

The SEM-R intervention includes three phases. In Phase 1, the "exposure" phase, teachers presented short read-alouds from high-quality literature focusing on high levels of cognitive engagement (Guthrie, Wigfield, & VonSecker, 2000; Knapp et al., 1995; Taylor, Pearson, Peterson, & Rodriguez, 2003) by selecting enjoyable "bookhooks" to introduce and expose students to a wide variety of titles, genres, authors, and topics (Renzulli, 1977; Renzulli & Reis, 1997). As part of these oral shared read-alouds, teachers provided scaffolded instruction through modeling and discussion, focusing on demonstrating reading strategies and self-regulation skills, such as those advocated in a more recently recommended type of scaffolded silent reading (Reutzel, Hollingsworth, & Eldredge, 1994; Reutzel, Jones, Fawson, & Smith, 2008) and the use of higher order questions to guide discussion (Taylor et al., 2003; Taylor, Pearson, Clark, & Walpole, 2000). Across all phases of SEM-R, researchers have implemented the ideas of teaching reading for cognitive engagement as discussed by Taylor et al. (2003), Knapp et al. (1995), and Guthrie, Wigfield, & VonSecker (2000).

Phase 2 of the SEM-R model emphasizes the development of students' ability to engage in supported independent reading (SIR) of self-selected, appropriately challenging books, with differentiated instruction in conferences with the teacher or another adult. Controversy about this topic ensued after the NRP reported a lack of research support for silent reading and discussed the shortcomings of both the report and some of the research examined in the report (see, e.g., Allington, 2002; Cunningham, 2001; Krashen, 2002). Research supporting various uses of independent reading during reading class exists. Duke (2000) found that students need extended experiences with print of various genres for continuing academic achievement. Anderson, Wilson, and Fielding (1988), studying the relation between the amount of student reading completed outside of school and reading achievement, identified reading books as the best predictor of reading achievement. Taylor et al. (1990) studied elementary students who kept daily reading logs, noting that time spent reading in school contributed to growth in reading achievement. Recently, Reutzel, Fawson, and Smith (2008) found that students who used silent, sustained reading did just as well as those who

used guided repeated oral reading on fluency and comprehension. Krashen (2002) studied the use of silent reading accompanied by some instruction using children's books, finding achievement benefits for independent reading. Some of the criticism leveled at silent reading specifies the shortcomings that exist when teachers do not give instruction or feedback during the process (Stahl, 2004). Conferences, accompanied by differentiated instruction during Phase 2 independent reading, are a core component of SEM-R, as is an emphasis on appropriate challenge levels of the books selected by students. Phase 2 includes most of the general principles of effective reading instruction identified by Reutzel and Smith (2004), including modeling and scaffolding during conferences, time on task, volume of reading, student choice, discussion and dialogue, access to a variety of reading materials, encouragement of engaged reading, a print-rich environment, and silent reading practice. During Phase 2, students selected books from a classroom library given as part of the study (Tivnan & Hemphill, 2005) that included high-quality, age-appropriate books (Teale & Gambrell, 2007) as well as high-interest and above-grade-level texts (Reis & Fogarty, 2006). Students were encouraged to select books that challenged them and were approximately one to two grade levels above their current independent reading levels, ensuring that they were of high interest and neither too easy nor too difficult (Allington & McGill-Franzen, 1989; Anderson et al., 1988, Reis & Renzulli, 1989). In the SEM-R, teachers monitored and evaluated book selection and assisted students in the selection of books of appropriate challenge during weekly conferences, as quantity and quality of book selections contribute to higher achievement (Topping, Samuels, & Paul, 2007). Another focus of book selection was related to nonfiction, as 30% to 40% of the classroom libraries across grade levels consisted of nonfiction books, shown to be effective in boosting comprehension, especially for boys (Topping, Samuels, & Paul, 2008). Students learned strategies for recognizing appropriate books and were coached to select challenging instructional-level books in areas of their interest to promote engagement (Guthrie & Wigfield, 2000; Wigfield & Guthrie, 1997). Over the course of the intervention, students initially read for 5 to 15 minutes a day during Phase 2; over time they extended SIR to 20 to 25 minutes and finally to 35 to 45 minutes each day.

During this in-class reading time, students participated in reading conferences with their teachers, receiving monitored and differentiated instruction (Bryan, Fawson, & Reutzel, 2003; Manning & Manning, 1984) during brief, individualized, instructional conferences. On average, each student participated in one to two conferences per week for a duration of about 5 minutes per conference. These conferences included time for positive interactions with students, a focus on differentiated student-centered instruction (Renzulli & Reis, 1997; Tomlinson, 2003), and supportive classroom interactions (Bryan et al., 2003; Manning & Manning, 1984; McAllister & Irvine, 2002). During student conferences, students read aloud brief sections of their

books (Hiebert, 2006), and teachers consistently monitored and documented the instructional challenge match of each book read in Phase 2 while also encouraging and praising students for success in reading (Thompson, Ransdell, & Rousseau, 2005). In these conferences, classroom teachers and instructional aides provided individualized instruction in strategy use, including predicting, using inferences, and making connections through modeling and discussions (Bandura, 1986; Bryan, et al., 2003; Dowhower, 1987; Duffy, Roehler, & Herrmann, 1988; Rasinski, 1990; Stahl, 2004). Phase 3 of SEM-R is also based on research that avoids ineffective reading instruction (Flippo, 1998), focusing instead on effective reading instruction as identified by experts in reading (Allington, 2001; Anderson et al., 1988; Snow, Burns, & Griffin, 1998) as well as enrichment and engagement pedagogy (Guthrie & Wigfield, 2000; Renzulli & Reis, 1985; Wigfield & Guthrie, 1997) to enable students to move from teacher-directed opportunities to self-choice activities over the course of the SEM-R study. Phase 3 activities included, but were not limited to, opportunities to explore new technology, discussion groups, practice with advanced questioning and thinking skills, creativity training in language arts (Renzulli, Callahan, Smith, Renzulli, & Gay, 2000), learning centers, Interest-based projects, buddy oral reading, and book discussions.

The following specific research questions guided the SEM-R research study:

1. Can the regular reading curriculum be replaced by an independent and interest-based program (SEM-R) without adversely affecting scores on standardized assessments of reading fluency and reading comprehension?
2. Can the use of the SEM-R increase students' reading fluency and comprehension?
3. Do teachers, principals, and literacy coaches of students who participate in the SEM-R intervention report higher levels of student engagement in reading?

Research Method

This study incorporated cluster-randomized assignment to groups, with 37 classrooms in the treatment condition and 33 in the control condition. The sample included students in second to fifth grades from five elementary schools across the United States. The schools were selected using criteria specified in the funding grant, with a focus on participation of schools in urban or rural settings with high percentages of students placed at risk because of poverty or other factors and the ability and willingness of school personnel to meet the methodological research requirements (e.g., random assignment to treatment or control condition for both teachers and students, integrated implementation of the treatment conditions, and timely administration of assessments). The SEM-R intervention for all five schools started within 2 weeks of the start of the fall academic school year and continued for 5 months, through the last 2 weeks of February. Pretest

and posttest data were collected on students' reading fluency and comprehension, and the quantitative procedures of hierarchical linear modeling (HLM) and multivariate analysis of variance were used to investigate the effects of the SEM-R intervention on these reading outcomes.

Sample

The SEM-R was implemented in five elementary schools, which were selected for participation in this study based on school size and demographics. Schools with at least three to preferably four classes per grade level across second to fifth grade were selected. The majority of classrooms were in third to fifth grade, but two second grade gifted classes reading at above grade level were also included. The number of participating teachers at each grade level for each school is summarized in Table 1. The total group of schools included rural, urban, and suburban locations across five states, and the student population at the five schools varied by race, ethnicity, and language as well as socioeconomic status (SES). In total, 1,192 students and 63 teachers participated in the study. Percentages of students within specific demographic groups at each school are summarized in Table 2 (school names are pseudonyms).

The teachers in the treatment and control conditions were similar in years of experience and highest degree attained (as reported on the preprogram Teaching and Reading: Attitudes and Practices Survey). Across the five schools, treatment group teachers had a mean of 13.8 years of experience (SD = 8.90) and control group teachers had a mean of 15.9 years of experience (SD = 11.04). Treatment teachers ranged from 3 to 33 years of experience, and control group teachers ranged from 1 to 37 years. Fourteen teachers from each group had achieved a bachelor's as their highest degree; 15 in the treatment group and 16 in the control group had achieved a master's degree. One teacher in the treatment group had achieved a 6th-year certificate.

An examination of years of experience between treatment group and control group teachers within schools showed that two of the schools had similar levels of experience between groups while the other three showed greater differences. In the Urban Magnet School, treatment teachers had a mean of 18.57 years of experience while control group teachers had a mean of 9.83 years. In both the Suburban South School and the Urban Southeast School, on the other hand, control group teachers had considerably more experience than treatment group teachers (25.67 years control, 11.63 years treatment in Suburban Southeast School, 12.50 years control, 6.67 years treatment in the Urban Southeast School).

Recruitment and Study Preparation Procedures

Recruitment of schools was completed in the year prior to the study through contacts to the researchers from previous articles published on

Table 1

Number of Participating Teachers by Grade Level and Condition

	Suburban South		Urban Southeast		Urban Magnet		Rural South		Suburban Midwest		
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Totals
2nd grade	1	1	0	0	0	0	0	0	0	0	2
3rd grade	4	3	1	2	3	2	3	3	2	2	25
4th grade	2	1	1	1	2	2	3	3	2	2	19
5th grade	1	1	1	1	2	2	2	2	2	2	16

Table 2
School Demographic Information

Category	Suburban South %	Urban Southeast %	Urban Magnet %	Rural South %	Suburban Midwest %
White	45	1	59	65	55
Black	37	93	15	28	14
Hispanic	10	5	6	3	9
Asian	2	1	20	1	16
American Indian	1	0	0	0	0
Multiracial	5	—	—	3	7
Disabled or special education	20	11	6	20	14
Limited English proficient	12	3	23	0	20
Gifted and talented (using district designation as such based on state criteria)	29	—	100	10	9
Economically disadvantaged, free or reduced meals	33	93	25	60	40

the SEM-R as well as conference presentations. Discussions of the study were held with interested principals, with emphasis on the need for random assignment to treatment and control groups and for a commitment to designating half of the daily 2-hour traditional language arts block to SEM-R in the treatment group. Institutional research board permission was sought and granted. The researchers conducted school visits with informal observations of reading classrooms to ensure that practices similar to the intervention were not already in use, and students and teachers were then randomly assigned to treatment or control groups. Each principal identified a research liaison to facilitate the implementation of the study, and one research team member was designated as a professional development coach for each school.

The next stage of preparation included 6 hours of professional development for all randomly selected SEM-R treatment teachers. Teachers received professional development during the summer before implementation and met the professional development coach from the research team who would be working with them throughout the intervention. They received written information about the SEM-R and a collection of approximately 250 books for their classroom library selected to include multiple levels of reading achievement across various interest areas. Attention was paid to the cultural background of students attending the schools, and input was also sought from literacy coaches about the books selected. A collection of bookmarks and student and teacher logs were also provided to each classroom. Professional development were given about challenge levels of the books,

with information about Fountas and Pinnell (2001) Guided Reading Level, Developmental Reading Assessment Level (MetaMetrics, 2004), and Lexile Levels (Scholastic, 2007) as well as the use of conferences and student read-alouds to determine appropriate complexity of text at slightly above grade level. At the end of the initial professional development session, classroom teachers were asked to begin to plan the implementation of the SEM-R in their classrooms for the following September.

As noted above, all treatment teachers initially received approximately 250 high-interest fiction and nonfiction books across several reading levels to support SEM-R implementation. Most teachers augmented their collection as the intervention continued, choosing literature based on students' interests and experiences from the school library or their own collections. Teachers also received sets of bookmarks that listed higher order questions; each bookmark listed three to five questions addressing a particular literary element, theme, genre, or other area of study. Teachers used the bookmarks in both Phase 1 discussions and Phase 2 conferences to promote higher order thinking.

Implementation Procedures

All schools in the study had a 2-hour block of each school day devoted to reading and language arts instruction. Teachers who were randomly assigned to the treatment group taught 1 hour of language arts, using the regular language arts program in the school, and taught SEM-R in the other hour of the block. In the 1 hour that was not devoted to SEM-R, teachers provided writing, vocabulary, and other spelling and language activities. SEM-R activities were documented in teacher and student logs, as teachers noted the activities conducted within each phase and students recorded the books they were reading, how long they spent reading each day, questions they had about their reading, and the reading strategies they had used that were helpful to them. The SEM-R represents a model for providing individualized reading instruction rather than a set curriculum. Thus, the implementation was standardized across grade levels; however, the actual instructional activities were grade level appropriate. Each grade level received a different classroom set of books; therefore, the reading and individualized conferencing was differentiated by grade level. Teachers randomly assigned to the control group continued providing instruction for 2 hours using the regular reading and language arts program across all grade levels. All schools selected had this 2-hour block for reading at all participating grade levels.

During the study period, school liaisons worked cooperatively with research team coaches to help with data collection and to provide teachers with assistance for implementation of the SEM-R. Research team members were available via email and phone during the intervention to provide

support and to monitor both intervention and control classrooms. Research team members made regular visits to the schools to conduct observations of treatment and control classes for treatment fidelity.

Instrumentation

Oral reading fluency. Measures of oral reading fluency (ORF) assess the speed, accuracy, and efficiency with which a student reads a particular text. Reading fluency was assessed before and after intervention in this study using procedures described by Hasbrouck and Tindal (2005). Reading passages were selected to represent all grade levels (3–5), and to facilitate comparisons across grade levels, all students read from the three increasingly difficult, 250-word passages for three separate 1-minute reading trials. The number of words read correctly for each passage was recorded, and a mean ORF score was calculated and recorded for each student. ORF reflects the complex orchestration of both lower level and higher level processes and is considered a reliable indicator of reading proficiency (Fuchs, Fuchs, Hosp, & Jenkins, 2001). Test-retest reliability from pre to post measures of ORF in this sample was .94, and the internal consistency reliability as determined by Cronbach's alpha for both pre and post fluency was .98.

Reading comprehension. Reading comprehension was measured before and after the intervention using the Iowa Tests of Basic Skills (ITBS) Reading Comprehension subtest (Form A). The ITBS measures achievement in 15 subject areas for students in Grades K–8. The Reading subtests of the ITBS "measure how students derive meaning from what they read" (Hoover et al., 2003, p. 32). The Reading Comprehension subtest consists of a variety of reading passages representing narrative, poetry, and nonfiction material from science and social studies. After students read each passage, their comprehension is assessed through the use of four to seven multiple-choice questions that ask students to recall facts, make generalizations, and draw inferences. For the Language Arts subscales of the ITBS, reliability coefficients are greater than .95 (Hoover et al., 2003). The ITBS is vertically scaled, thus students' scores on the different forms of the ITBS were comparable across grade levels.

Teaching and Reading: Attitudes and Practices Survey (TRAPS). Treatment and control teachers completed the TRAPS (Fogarty, Little, & Reis, 2005) at the beginning and at the end of the study. This instrument drew on previous research demonstrating a significant linear relationship between teachers' own engagement in personal reading and their use of recommended practices in classrooms (Morrison, Jacobs, & Swinyard, 1999); some items were modeled, with permission, on items from the Morrison et al. (1999) survey. The TRAPS includes (a) a section requesting

demographic details from teachers, including years of experience, gender, and highest earned degree; (b) a section assessing teacher attitudes toward personal reading and the teaching of reading; and (c) a section asking a variety of questions about teachers' current practices in reading. For purposes of this study, data were examined to determine the degree to which treatment and control group teachers were similar at the beginning of the study in experience and attitudes. Teacher experience data are reported in the sample section.

Data Collection Procedures

Reading achievement tests. During the first weeks of the SEM-R intervention, a team of researchers traveled to the participating schools to collect pretest data. The ITBS was administered by treatment and control teachers in their respective classrooms under the supervision of the SEM-R research team in each school. ORF pretests were administered by the research team and individuals trained by the research team members in all five schools during the course of 2 to 3 days at each school. On the days of the ORF assessments, procedures were reviewed and each test administrator was observed to ensure that the assessments were standardized. Test administrators were given a stopwatch, a copy of each of the three reading passages to be reused for each student, and a second copy of the same passages stapled together with a blank student data sheet. The test administrators began each assessment by introducing themselves, recording the student's name, grade, and current reading teacher on the blank student data sheet, and then inviting the student to read.

The same procedures for both ITBS and ORF testing were followed for collecting posttest data in the final weeks of the study. All ORF score sheets and score averages were checked by two members of the research team for accuracy in calculations.

Observations for treatment fidelity in SEM-R classrooms. Research team members regularly monitored treatment fidelity in the schools during the intervention. Observations were conducted in treatment classes and control classes, with an average of one to two observations per classroom per month of the study. During 250 separate observations in treatment classes, researchers took detailed notes on the specific features of each phase of the SEM-R, using the SEM-R Observation Scale (Little, Fogarty, & Reis, 2005) and completing field notes. The SEM-R Observation Scale (for treatment classes) includes a nine-item fidelity form on which observers indicate whether or not particular SEM-R elements were present during the observation (see Table 3). The scale also includes space for rating whether the quality of particular elements was below, at, or above expectations, based on specific descriptors, and space for recording detailed field notes about the observed session, with guidelines for specific features to note by phase; for purposes of the

Table 3
Treatment Fidelity Checklist for the Schoolwide Enrichment Model–Reading

Phase	Key Criteria
Phase 1	1. Provided exposure by introducing books with a book discussion
	2. Read aloud from books that appeared to be selected in advance
	3. Integrated reading strategies and/or higher level thinking questions (e.g., bookmark questions)
Phase 2	4. Provided time for students' supported independent reading
	5. Established an environment in which students utilized self-regulation for supported, independent reading time
	6. Listened to students read in individual conferences
	7. Provided differentiated reading strategies and/or literary discussions during conferences
Phase 3	8. Made time available for Phase 3 independent or small group enrichment choices
	9. Provided 3–4 choices for students such as creativity training, Renzulli Learning, opportunities for individual reading, buddy reading, and other choices

present study, the nine-item fidelity form was used to document treatment fidelity quantitatively. To ensure interrater reliability, all observers received training on the use of the form, and selected observations were conducted by multiple observers and checked for reliability of ratings. Across 15 sessions that were observed by multiple team members, 9 instances demonstrated 100% agreement on ratings for the nine-item fidelity sheet; observers for 5 of the remaining sessions reached 89% agreement (eight of nine items). In the remaining case, only four of the items were expected to be observed based on scheduling realities; observers agreed on three of four items (75%). Therefore, interrater agreement was high on the observation scale.

Data from the nine-item fidelity sheet were reviewed and summarized by teacher and by school to assess fidelity of implementation. For any observed session, observers noted whether each item was observed or not observed; a rating of "not observed" carried the implication that an item was *expected* to be observed based on the phases being presented. In certain cases, a full SEM-R session might have been devoted to Phases 1 and 2, with no Phase 3, or observers might have entered after Phase 1 had already concluded. In these cases, the phases that were not observed were not rated. For each observation, a percentage score was calculated representing items observed as a percentage of items expected to be observed (as Phase 3 normally was scheduled once per week); each teacher's percentages were then averaged across all of his or her observations to reach a fidelity score.

Table 4
Fidelity of Implementation Results for Treatment Classes by School

School	*n* (Treatment Teachers)	Percentage Scores of Schoolwide Enrichment Model–Reading Elements Observed Across Observations[a]		
		Min	Max	*Mdn*
Suburban South	8	79	98	89
Urban Southeast	3	84	96	92
Urban Magnet	7	76	100	86
Southern Rural	8	61	99	86
Midwest Suburban	6	76	100	100

[a]Percentage scores for each teacher calculated as items observed divided by items expected within each observation session, then averaged across all observations.

Overall, fidelity scores for treatment teachers were high, with most teachers achieving average percentages above 80%; a careful review of field notes from the remaining portions of the observation scale supported these inferences regarding fidelity of implementation of SEM-R. Table 4 summarizes the percentages of observed SEM-R elements by school.

Observation of control classrooms. The observation scale used in control classrooms asked observers to note what types of reading activities (e.g., teacher reads aloud, students read aloud, guided reading, workbook exercises, etc.) occurred within each 10- to 15-minute segment of the observed session and to indicate the types of materials and instructional grouping strategies used. Observers also wrote specific notes describing the details of what they observed in the classrooms.

In the control classes, observers completed the SEM-R Observation Scale for Control Classrooms that included detailed field notes throughout the observed sessions. On the observation scale, trained observers indicated the type of reading activity from a number of selections (reading instruction, teacher read aloud, student read aloud, student silent reading, literacy activities, workbook exercises, and other activities, such as test preparation. During the course of the intervention, 161 control classroom observations were completed by observers trained in recording reading activities, with the most frequent activities indicated in descending order. Reading instruction, most frequently identified as teacher-led work and reading from basal books or class novels, was noted as 46% of all recorded activities. The next most frequently used strategy was silent reading, noted as 31% of all recorded activities and defined as reading from student-selected trade books or multiple copies of the same book with no teacher monitoring or instruction. The third most frequent activity fell under the category of other

activities and included test preparation activities such as explicit practice for state or local tests and work on vocabulary or writing tasks for the state or local assessments given, and this was noted 24% of the time. The fourth most frequently observed activity was workbook exercises using basal textbooks and their accompanying workbooks, noted 21%. One of the most consistent findings about the control classrooms was that the vast majority of all instruction was delivered to the whole classroom at one time, and this occurred 70% of time. Across all observations, the students were broken down into small groups or given individual instruction 21% of the time.

In summary, our observations of control classrooms confirmed that instruction was primarily based on the basal reading program or novel study that had been identified as the reading program used at each school. The reading programs as part of the language arts curriculum used in the control schools varied but included basal programs with newer and older versions of Scott Foresman and Accelerated Reader used in combination with an older basal program, remedial reading programs used in combination with district test preparation workbooks delivered monthly, and combinations of class novels, used in combination with older basal programs. A fairly consistent set of activities occurred each day, including whole group instruction, some smaller group reading instruction, and 3 to 5 minutes of reading followed by workbook and other activities. In all schools, in addition to the varied activities from the basal reading series, students participated in a wide variety of learning activities as summarized in the data earlier.

Qualitative Data Collection and Analysis

Site visits made to each school over the 5-month intervention included classroom observations as well as a review of teacher and student logs and researchers' site visit observations. These data, the accompanying interview data, and data from the SEM-R Observation Scale from classroom observations were used to triangulate sources. Across the schools, each principal was interviewed, as were each of the five SEM-R literacy coaches and all 32 SEM-R teachers. An interview protocol was used with common questions asked of all teachers, literacy coaches, and principals. The majority of the questions related to perceived benefits of the use of SEM-R in general and for students and teachers as well. A second set of questions related to perceptions of changes in student reading.

In addition, 250 SEM-R class observations enabled researchers to identify the details, contexts, and patterns of reading instruction across the SEM-R treatment classrooms using the SEM-R Observation Scale for each site and coding for benefits to students. The major benefits were indicators of student engagement in reading. Observations included a systematic description of events and student behaviors during SEM-R classrooms (Yin, 2002). Student engagement in reading was documented by the number of minutes

students read as well as the number of students engaged in independent reading across the observation period. This amount of time was necessary to compare outcomes across several schools and to triangulate data across observations, checklists, and interviews (Marshall & Rossman, 1989; Miles & Huberman, 1994). Data were coded manually using Strauss and Corbin's (1990) data coding paradigm and verified using metamatrices, master charts that organize data from each of the cases into a standard format (Miles & Huberman, 1994) to enable patterns and themes to emerge. Data were coded, as suggested by the Strauss and Corbin coding, into three levels of coding techniques—open coding, axial coding, and selective coding. Three researchers independently coded and then conferred with each other to confirm the decisions made about initial coding and emerging categories. For example, open coding responses to interview questions about student benefits from the use of SEM-R included responses such as "I have never had so many of my students fully engrossed in reading," "My students love to read now," "By far, the greatest change is student engagement," "Students loving to read is the greatest change in using SEM-R," and "Students reading more challenging books has increased their engagement." Each of these open codes were entered on a metamatrix, and then axial codes were determined relating to perceptions of benefits and changes. Finally, the core category was identified, which was positive changes in student engagement because of the use of SEM-R as discussed in the results section.

Data Analysis

To examine the differences in reading fluency and comprehension across instructional groups at the five elementary schools, we conducted a series of multilevel models using HLM version 6.04 (Raudenbush, Bryk, Cheong, Congdon, & Toit, 2004). The outcome variables included in the analyses were the postassessment scores on ORF measures and on the Reading Comprehension section of the ITBS. The correlation between the fluency measure and the ITBS (at posttest) was .74. Ethnicity, gender, and preassessment scores served as student-level covariates. Tables 5 and 6 include the fluency and ITBS assessment descriptive statistics for the pretests and posttests. Given the small Level 2 sample size (J = 70 classrooms), we used restricted maximum likelihood estimation (Raudenbush & Bryk, 2002). However, to conduct any necessary model fit tests, we reran the analyses using full maximum likelihood estimation. This allowed us to compare the deviances of nested models using the chi-square difference test and to compute model fit statistics such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC).

The Level 1 model contained three student-level variables: pretest scores, gender, and ethnicity. Gender was coded as 0 = *male* and 1 = *female*.

Table 5
Fluency Descriptive Statistics for Treatment and Control Groups in Five Elementary Schools

School and Grade	Treatment Group					Control Group				
	Prefluency		Postfluency			Prefluency		Postfluency		
	M	*SD*	*M*	*SD*	*n*	*M*	*SD*	*M*	*SD*	*n*
Suburban South										
2nd grade	96.15	28.38	123.22	27.15	20	88.62	22.71	114.17	21.19	23
3rd grade	96.72	41.16	114.58	40.18	70	106.96	46.05	119.82	43.96	47
4th grade	147.81	30.47	167.55	30.77	38	159.95	23.34	177.59	19.68	21
5th grade	172.80	24.84	192.38	25.52	20	161.00	33.16	175.04	28.57	15
Urban Magnet										
3rd grade	128.56	39.81	149.27	35.97	68	119.95	40.09	133.49	39.85	49
4th grade	145.47	37.78	164.08	35.63	52	147.01	34.48	172.16	32.25	53
5th grade	166.87	39.82	185.45	39.02	55	158.11	38.65	172.70	42.17	59
Midwest Suburban										
3rd grade	75.22	31.48	100.46	32.01	37	90.02	39.77	118.29	42.07	38
4th grade	114.98	38.13	134.14	42.76	35	119.98	41.40	142.03	42.54	33
5th grade	120.13	37.19	139.01	41.86	36	125.96	32.04	144.27	34.90	38
Southern Rural										
3rd grade	76.25	37.69	98.26	38.81	39	80.31	37.10	109.20	41.64	31
4th grade	95.33	41.26	109.66	38.62	45	99.36	39.65	117.33	39.02	29
5th grade	123.29	37.68	140.03	37.67	52	119.19	38.76	139.98	34.23	31
Urban Southeast										
3rd grade	63.17	29.82	93.19	35.70	25	48.00	23.12	65.00	27.47	19
4th grade	81.38	32.95	110.71	37.24	35	76.82	28.84	82.78	31.13	35
5th grade	104.83	38.59	119.18	46.71	22	81.65	24.20	96.52	25.56	22

Ethnicity was dummy coded so that underrepresented minority students (which included African American or Black, Latino/a, and Native American) were coded as 1. Students of all other racial/ethnic categories (i.e., White, Asian, and Other) served as the reference category. The Level 2 model contained both classroom and school-level information. Level 2 variables included treatment group, school dummy variables, and school by treatment interaction terms. The SEM-R treatment group was coded as 1, and the control group was coded as 0. The five schools were treated as fixed effects at Level 2. We created a series of four dummy codes to represent the five schools. In this coding system, Urban Southeast served as the reference group. We also created four school by treatment product terms to model potential interactions between the school and treatment variables. For all analyses, we grand mean centered pretest scores; we left all categorical

Table 6
Iowa Tests of Basic Skills Descriptive Statistics for Treatment and Control Groups in Five Elementary Schools

School and Grade	Treatment Group					Control Group				
	Pre-ITBS		Post-ITBS			Pre-ITBS		Post-ITBS		
	M	*SD*	*M*	*SD*	*n*	*M*	*SD*	*M*	*SD*	*n*
Suburban South										
2nd grade	179.20	13.32	196.25	17.97	20	183.78	10.21	197.09	15.69	23
3rd grade	180.56	24.02	198.30	26.11	70	183.81	24.13	205.91	24.63	47
4th grade	221.97	17.65	233.03	20.76	38	229.62	14.43	247.10	14.75	21
5th grade	233.75	14.90	256.15	18.11	20	238.13	17.09	253.47	23.69	15
Urban Magnet										
3rd grade	201.78	24.05	212.68	24.40	68	198.45	22.86	215.51	26.18	49
4th grade	220.15	27.38	238.42	29.21	52	225.74	23.05	241.55	21.41	53
5th grade	239.96	29.75	254.69	26.13	55	236.44	31.09	247.41	29.80	59
Midwest Suburban										
3rd grade	177.14	21.54	194.78	23.14	37	180.74	20.37	199.18	24.94	38
4th grade	200.17	28.99	216.46	26.89	35	200.58	23.51	213.70	20.69	33
5th grade	212.17	26.18	228.83	31.82	36	208.29	27.84	215.18	29.14	38
Southern Rural										
3rd grade	177.87	21.22	195.54	21.37	39	174.61	23.90	192.13	24.61	31
4th grade	193.33	31.27	201.60	30.01	45	203.14	28.23	209.86	26.18	29
5th grade	209.56	23.35	220.08	29.97	52	209.58	27.32	223.13	23.52	31
Urban Southeast										
3rd grade	164.04	17.47	172.08	20.47	25	153.53	9.35	158.79	10.64	19
4th grade	173.57	17.95	188.80	22.53	35	166.97	17.12	172.89	15.56	35
5th grade	188.45	21.69	191.50	29.66	22	174.59	15.92	178.73	17.95	22

Note. ITBS = Iowa Tests of Basic Skills.

variables uncentered. Finally, we included three dummy-coded grade level variables at Level 2, and we chose fifth grade as the reference group.

Model for Reading Fluency

Three alternative models were investigated before we determined our final model for the analysis of reading fluency. Each of the three models contained schools, treatment, and the school by treatment interaction terms as predictors of the Level 1 intercept. However, the three models differed in terms of which Level 1 and Level 2 covariates were included in the model. The first model included only one Level 1 covariate: pretest scores for reading fluency. At Level 2, schools (the four dummy coded variables), treatment, and the school by treatment interaction terms were predictors of the Level 1 intercept. The second model included three Level 1 covariates: pretest scores

Table 7
Model Comparison Table for Reading Fluency Results

Model	Deviance	Parameters	Akaike Information Criterion	Bayesian Information Criterion
Model 1	9924.47	15	9954.17	9988.20
Model 2	9909.39	20	9949.39	9994.36
Model 3	9911.12	17	9945.12	9983.35

for reading fluency, gender, and ethnicity. In addition to including schools, treatment, and the school by treatment interaction terms, Model 2 included grade level (three dummy coded variables) to predict the Level 1 intercept. Finally, our third model included gender, ethnicity, and pretest scores for reading fluency as Level 1 covariates and schools, treatment, and the school by treatment interaction as Level 2 predictors of the intercept. We found that after controlling for reading fluency, grade level failed to explain any additional variance in reading fluency scores. In addition, the chi-square difference test, the AIC, and the BIC favored Model 3, so Model 3 became our final model of reading fluency. Table 7 contains the model fit statistics for all three models of reading fluency.

The final model for reading fluency appears below:

Level 1 model:

$$Y_{ij} = \beta_{0j} + \beta_{1j}\,(\text{pretest fluency}) + \beta_{2j}\,(\text{gender}) + \beta_{1j}\,(\text{ethnicity}) + r_{ij}$$

Level 2 model:

$$\begin{aligned}\beta_{0j} = {} & \gamma_{00} + \gamma_{01}\,(\text{Suburban South}) + \gamma_{02}\,(\text{Suburban South} \times \text{treatment}) \\ & + \gamma_{03}\,(\text{Urban Magnet}) + \gamma_{04}\,(\text{Urban Magnet} \times \text{treatment}) \\ & + \gamma_{05}\,(\text{treatment}) + \gamma_{06}\,(\text{Midwest Suburban}) \\ & + \gamma_{07}\,(\text{Midwest Suburban} \times \text{treatment}) + \gamma_{08}\,(\text{Southern Rural}) \\ & + \gamma_{09}\,(\text{Southern Rural} \times \text{treatment}) + u_{0j}\end{aligned}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

Model for Reading Comprehension

Before determining our final model for reading comprehension, we investigated four alternative models. The first model included the pretest for reading comprehension as the sole Level 1 covariate. Level 2 variables included schools, treatment, and the school by treatment interaction terms

as predictors of the Level 1 intercept. The second model included three student-level covariates: the reading comprehension pretest score, gender, and ethnicity. The Level 2 variables included schools, treatment, and the school by treatment interaction terms as predictors of the Level 1 intercept. The third model included gender, ethnicity, and the pretest score at Level 1 and schools, treatment, the school by treatment interaction terms, and grade level as Level 2 predictors of the Level 1 intercept. The fourth model included the same Level 1 variables as the third model, and it included the same Level 2 predictors of the intercept. However, the Level 2 variables (schools, treatment, the school by treatment interaction terms, and grade level) were also included as Level 2 predictors for the pretest slope. Table 8 contains the model fit statistics for all four models for reading comprehension. The chi-square difference test, the AIC, and the BIC all indicated that the third model provided the best fit to the data. Therefore, Model 3 became our final model for reading comprehension.

The final model for reading comprehension appears below:

Level 1 model:

$$Y_{ij} = \beta_{0j} + \beta_{1j}\,(\text{pretest comprehension}) + \beta_{2j}\,(\text{gender}) + \beta_{1j}\,(\text{ethnicity}) + r_{ij}$$

Level 2 model:

$$\begin{aligned}\beta_{0j} = {} & \gamma_{00} + \gamma_{01}\,(\text{Suburban South}) + \gamma_{02}\,(\text{Suburban South} \times \text{treatment}) \\ & + \gamma_{03}\,(\text{Urban Magnet}) + \gamma_{04}\,(\text{Urban Magnet} \times \text{treatment}) \\ & + \gamma_{05}\,(\text{treatment}) + \gamma_{06}\,(\text{Midwest Suburban}) \\ & + \gamma_{07}\,(\text{Midwest Suburban} \times \text{treatment}) + \gamma_{08}\,(\text{Southern Rural}) \\ & + \gamma_{09}\,(\text{Southern Rural} \times \text{treatment}) + \gamma_{010}\,(\text{Second grade}) \\ & + \gamma_{011}\,(\text{third grade}) + \gamma_{012}\,(\text{fourth}) + u_{0j}\end{aligned}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

We examined standard statistical assumptions of HLM including the normality of Level 1 residuals and homogeneity of Level 1 variances. The Level 1 residuals for fluency and comprehension were normally distributed. No major violations were noted for the test of homogeneity of Level 1 variance. The Level 1 variables were independent of the Level 1 residuals. In addition, at Level 2, the empirical Bayes residuals for the intercept and slope for the dependent variables were normally distributed, and they were uncorrelated with all of the Level 2 predictors.

Table 8
Model Comparison Table for Reading Comprehension Results

Model	Deviance	Parameters	Akaike Information Criterion	Bayesian Information Criterion
Model 1	10311.85	15	10341.85	10405.57
Model 2	10298.20	17	10332.20	10370.43
Model 3	10280.87	20	10320.87	10365.83
Model 4	10259.53	32	10323.53	10394.95

Results

Reading Fluency

First, to estimate the intraclass correlation (ICC), we estimated an unconditional random effects ANOVA model with no predictors at Level 1 or Level 2. The ICC, which measures the proportion of variance in reading fluency between classrooms, is equal to the between-classroom variance (τ_{00}) divided by the total variance ($\tau_{00} + \sigma^2$). The ICC for reading fluency was .426. In other words, 42.6% of the variance in reading fluency was between classrooms whereas 57.4% of the total variability was within classrooms. This between-class variation was statistically significant, $\tau_{00} = 937.06$, $\chi^2(69) = 990.32$, $p < .001$.

Second, we estimated a Level 1 model that included pretest scores for reading fluency, ethnicity, and gender as predictors of posttest scores for reading fluency, but that did not include any Level 2 predictors. The addition of these Level 1 predictors reduced the within-school variability in reading fluency by 82.17%. In other words, pretest fluency scores, gender, and ethnicity accounted for about 82.17% of the student-level variance in the posttest scores in the intercept. Furthermore, the addition of pretest fluency scores, gender, and ethnicity explained 96.6% of the between-class variance in the intercept for posttest reading fluency. However, there was still statistically significant between-class residual variance in the intercept, $\tau_{00} = 31.52$, $\chi^2(69) = 208.02$, $p < .001$, remaining to be explained. The slope of pretest reading fluency scores on posttest reading fluency scores did vary across classrooms; therefore, we estimated a random effect for the prefluency slope. Finally, we estimated the full Level 2 model (Model 3) discussed previously. The Level 2 model explained an additional 10.82% ((31.52 – 28.11) / 31.52) of the between-classroom variance in the intercept relative to the random coefficient model. After accounting for the variables at Level 1 and Level 2, there was statistically significant between-class variability remaining to be explained in both the intercept, $\chi^2(60) = 183.46$, $p < .001$, and slope of the pretest score for reading fluency, $\chi^2(69) = 132.99$, $p < .001$.

Given the coding scheme used in this analysis, the overall intercept, γ_{00} = 126.42, $t(60)$ = 34.37, $p < .001$, represents the predicted posttest fluency score for a nonminority male student with an average pretest score for reading fluency (pretest was grand mean centered) who attended Urban Southeast Elementary School (the reference group) and who was in the control group (SEM-R = 0). Therefore, the coefficient for Suburban South represents the differential between Suburban South and Urban Southeast for the comparison group after controlling for ethnicity, pretest for reading fluency, and gender. The average score for a similar student at Suburban South was 131.19 (126.42 + 4.77); however, this difference was not statistically significant. There was a statistically significant difference on the posttest reading fluency score between control group students at Urban Magnet and at Urban Southeast Elementary after controlling for pretest fluency scores, gender, and ethnicity, γ_{03} = 8.62, $t(60)$ = 2.05, p = .045. The average posttest reading fluency score for the control group at Urban Magnet was 135.04 (126.42 + 8.62) after controlling their pretest reading fluency scores, gender, and ethnicity. The control group students attending Midwest Suburban had predicted posttest reading fluency scores that were 12.21 points higher, γ_{06} = 12.21, $t(60)$ = 2.79, p = .007, and control group students attending at Southern Rural had predicted scores that were 9.7 points higher, γ_{08} = 9.70, $t(60)$ = 2.21, p = .031, than those of their peers at Urban Southeast, after controlling for pretest fluency scores, ethnicity, and gender. Table 9 compares the results of the baseline model, the random coefficients model, and the final contextual model.

The coefficient for interaction term between school and treatment represents the differential effect of SEM-R treatment at the named school as compared to Urban Southeast Elementary, which was the reference school. The coefficient for interaction term between Suburban South and treatment was not statistically significantly different from 0. The coefficients for interaction terms between each of the other schools and treatment were statistically significantly different, as follows: Urban Magnet by treatment, γ_{04} = –14.52, $t(60)$ = –2.67, p = .01; Midwest Suburban by treatment, γ_{07} = –17.47, $t(60)$ = –2.97, p = .05; and Southern Rural by treatment, γ_{07} = –20.91, $t(60)$ = –3.64, p = .01. Therefore, the effectiveness of the SEM-R treatment varied from school to school. SEM-R appeared to be most effective at Urban Southeast. Table 10 summarizes the predicted posttest reading fluency scores for each school.

The coefficient for SEM-R represents the predicted posttest reading fluency differential between a student at Urban Southeast who received the SEM-R treatment and a control student, after controlling for gender, fluency pretest, and ethnicity. Thus, the predicted posttest fluency score in the treatment group at Urban Southeast for a student of an average pretest reading fluency score, after controlling gender and ethnicity, was 141.73 (126.42 + 15.31), and this differential was statistically significant, γ_{05} = 15.31, $t(60)$ =

Table 9
Hierarchical Linear Modeling Results for Reading Fluency

Parameter	Unconditional Model		Level 1 Model		Full Model	
	Parameter Estimate	*SE*	Parameter Estimate	*SE*	Parameter Estimate	*SE*
Fixed effect						
Intercept (γ_{00})	129.03***	3.82	134.76***	1.02	126.42***	3.68
Suburban South (γ_{01})					4.77	4.31
Urban Magnet (γ_{03})					8.62*	4.21
Midwest Suburban (γ_{06})					12.21**	4.37
Southern Rural (γ_{08})					9.70*	4.39
SEM-R (γ_{05})					15.31**	4.67
Suburban South × SEM-R (γ_{02})					–10.47	5.66
Urban Magnet × SEM-R (γ_{04})					–14.52**	5.44
Midwest Suburban × SEM-R (γ_{07})					–17.48**	5.89
Southern Rural × SEM-R (γ_{09})					–20.91***	5.75
Prefluency (γ_{10})			0.91***	0.01		
Gender (γ_{20})			2.35**	0.91		
Represented (γ_{30})			–2.86*	1.08		
Variance estimate						
Level 1 variance (σ^2)	1263.001	35.54	225.150	15.00	225.146	
Intercept variance (τ_{00})	937.05	30.61	31.520	5.61	28.106	
Slope variance (τ_{11})			0.005	0.073	0.005	
Other statistics						
Deviance	12047.10		9933.87		9876.95	

Note. SEM-R = Schoolwide Enrichment Model–Reading.
*$p < .05$. **$p < .01$. ***$p < .001$.

3.28, $p = .002$. However, this coefficient represents the treatment effect at the reference school not an overall treatment effect. To determine whether the treatment effects were statistically significant at the other four schools, we reran the analyses, using each of the other four schools as the reference group. The results of these analyses revealed that treatment effects for Urban Magnet, Midwest Suburban, and Southern Rural were not statistically significant. However, the difference between control group and treatment group at Suburban South was statistically significant, SEM-R = 4.84, $t(60) = 2.21$, $p = .031$. Therefore, it is inappropriate to infer an overall treatment effect from these results. Instead, SEM-R appeared to affect fluency results differentially across schools. Table 11 shows the effect sizes for the difference between the treatment and control groups for each of the five schools in Cohen's *d* units. The standardized mean differences ranged from a low of

Table 10

Predicted Posttest Reading Fluency Scores of Students Whose Pretest Reading Fluency Scores Were at the Grand Mean After Controlling for Their Gender and Ethnicity

	Suburban South	Urban Magnet	Midwest Suburban	Southern Rural	Urban Southeast
Treatment	136.03	135.83	136.46	130.52	141.73
Control	131.19	135.04	138.63	136.12	126.42
Difference	4.84	0.79	–2.17	–5.6	15.31

Table 11

Differences Between the Means of the Treatment and Control Groups for Each School in Standard Deviation Units (Cohen's *d*)

School	Fluency Effect Size (*d*)	Comprehension Effect Size (*d*)
Suburban South	.10	–.11
Urban Magnet	.02	–.03
Suburban Midwest	–.05	.11
Southern Rural	–.12	–.01
Urban Southeast	.33	.27

–0.05 at Midwest Suburban to a high of 0.33 at Urban Southeast, suggesting that there were appreciable differences between the fluency scores of the treatment and control groups at some schools but not all of them.

The intercept for the pretest reading fluency slope, $\gamma_{10} = 0.91$, $t(69) = 60.53$, $p < .001$, represents the average effect of a Urban Southeast control group student's pretest reading fluency score on his or her posttest reading fluency score, after controlling for gender and ethnicity. For students in the control group at Urban Southeast, after controlling for gender and ethnicity, for every unit increase in the students' pretest fluency scores, their predicted posttest reading fluency scores increased by 0.91 units. Females' posttest reading fluency scores were 2.32 units higher than male students' scores, after controlling for their pretest reading fluency score and ethnicity. Also, underrepresented students (Hispanic, African American, and Native American students) had posttest reading fluency scores that were 2.92 points lower than those of White, Asian, and other students, holding school, treatment, pretest fluency scores, and gender constant.

Reading Comprehension (ITBS)

First, to estimate the ICC, we fit an unconditional random effects ANOVA model. The ICC for reading comprehension was .491. In other words, 49.1% of the variance in reading comprehension was between-classrooms

variance, whereas 50.9% was within classrooms. There was significant variation in the mean for reading comprehension, $\tau_{00} = 558.59$, $\chi^2(69) = 1269.95$, $p < .001$. Table 12 shows the results of all models, including the baseline models.

Second, we estimated a Level 1 model that included pretest reading comprehension score, gender, and ethnicity as predictors of posttest reading comprehension score. The addition of these predictors at Level 1 reduced the within-classroom variance in reading comprehension by 45.50%. Furthermore, the addition of these Level 1 variables reduced the between-class variance in reading comprehension by 91.4%. However, there was still statistically significant between-classroom residual variance in the intercept, $\tau_{00} = 48.07$, $\chi^2(69) = 195.28$, $p < .001$. Also, there was statistically significant variance between classrooms in the effect of pretest comprehension scores on posttest comprehension scores, $\tau_{11} = 0.02$, $\chi^2(69) = 127.47$, $p < .001$.

Finally, we estimated the full Level 2 model (Model 3) discussed previously. The Level 2 model resulted in a 61.68% reduction in between-class variance in the intercept relative to the random coefficient model. After accounting for the variables both at Level 1 and Level 2, there was statistically significant variability remaining to be explained in the intercept, $\tau_{00} = 18.42$, $\chi^2(57) = 128.88$, $p < .001$, and in the pretest reading comprehension slope, $\tau_{11} = 0.01$, $\chi^2(69) = 115.30$, $p = .001$.

Given the coding scheme used in this analysis, the intercept, $\gamma_{00} = 198.00$, $t(57) = 57.56$, $p < .001$, represents the predicted posttest comprehension score for a nonminority male student with an average pretest reading comprehension score attending Urban Southeast (the reference group) who was in the control group (SEM-R = 0). Therefore, the coefficient for Suburban South, $\gamma_{01} = 26.51$, $t(57) = 6.35$, $p < .001$, represents the predicted differential between control group students from Suburban South and Urban Southeast. Thus, the model-predicted score for a fifth-grade nonminority male control group student with an average pretest score for reading comprehension at Suburban South was 224.51 (198.00 + 26.51).

After controlling for gender and ethnicity, for every point higher a student scored on the reading comprehension pretest, he or she would be expected to score 0.69 points higher on the posttest, $\gamma_{10} = 0.69$, $t(69) = 25.81$, $p < .001$. After controlling for all other variables in the model, females did better than males, $\gamma_{20} = 3.16$, $t(1174) = 2.98$, $p = .003$. White and Asian students outperformed other students after controlling for all the other variables in the model, $\gamma_{30} = -3.30$, $t(1174) = -2.45$, $p = .015$.

At Urban Magnet, the model-predicted posttest score for a fifth grade nonminority student in the control group with a mean pretest score was 223.32 (198.00 + 25.32). At Midwest Suburban, the model-predicted score for a fifth grade nonminority student in the control group with a mean pretest score was 213.98 (198.00 + 15.98). At Southern Rural, the model predicted score for a fifth grade nonminority student in the control group

Table 12
Hierarchical Linear Modeling Results for Reading Comprehension

	Unconditional Model		Level 1 Model		Full Model	
Parameter	Parameter Estimate	*SE*	Parameter Estimate	*SE*	Parameter Estimate	*SE*
Fixed effect						
Intercept (γ_{00})	209.62***	2.92	213.54***	1.26	198.00***	3.44
Suburban South (γ_{01})					26.51***	4.17
Urban Magnet (γ_{03})					25.32***	4.01
Midwest Suburban (γ_{06})					15.98***	3.95
Southern Rural (γ_{08})					15.63***	3.95
SEM-R (γ_{05})					8.95*	3.99
Suburban South × SEM-R (γ_{02})					–12.72*	5.26
Urban Magnet × SEM-R (γ_{04})					–9.82	5.20
Midwest Suburban × SEM-R (γ_{07})					–5.14	5.31
Southern Rural × SEM-R (γ_{09})					–9.16	5.15
Second grade (γ_{010})					–13.71**	4.85
Third grade (γ_{011})					–6.91**	2.11
Fourth grade (γ_{012})					–1.72	2.04
Pre-ITBS (γ_{10})			0.75***	0.03	0.69***	0.03
Gender (γ_{20})			2.94**	1.07	3.16**	1.06
Represented (γ_{30})			–4.45***	1.31	–3.30*	1.35
Variance estimate						
Level 1 variance (σ^2)	579.63	24.08	315.91	17.77	316.84	17.80
Intercept variance (τ_{00})	558.59***	23.63	48.07***	6.93	18.42***	4.29
Slope variance (τ_{11})			0.02***	0.13	0.01***	0.1
Error covariance(τ_{10})			0.01	0.82	0.09	0.19
Other statistics						
Deviance (number of REML parameters)	11137.76		10341.04		10236.70	
Intraclass correlation	.4908					

Note. SEM-R = Schoolwide Enrichment Model–Reading; ITBS = Iowa Tests of Basic Skills; REML = restricted maximum likelihood estimation.
*$p < .05$. **$p < .01$. ***$p < .001$.

with a mean pretest score was 213.63 (198.00 + 15.63). The coefficient for second grade (γ_{010} = 13.71, $p < .001$) represents the differential between the predicted score for a second grade student and that of a fifth grade student after controlling for precomprehension (and other variables in the model). In other words, after controlling for all other variables in the model,

Table 13
Predicted Posttest Reading Comprehension Scores Whose Pretest Reading Comprehension Scores Were at the Grand Mean After Controlling for Their Gender, Ethnicity, and Grade Level

5th Grade	Suburban South	Urban Magnet	Midwest Suburban	Southern Rural	Urban Southeast
Treatment	220.74	222.45	217.79	213.42	206.95
Control	224.51	223.32	213.98	213.63	198.00
Difference	–3.77	–0.87	3.81	–0.21	8.95

second graders' reading fluency scores were, on average, 13.71 points lower than those of fifth grade students.

The coefficient for interaction term between Suburban South and treatment represents the differential effect of SEM-R treatment at Suburban South compared to that at Urban Southeast. Therefore, the predicted posttest reading comprehension score for a fifth grade male nonminority student of average pretest comprehension score who was exposed to SEM-R treatment at Suburban South Elementary was 220.74 (198.00 + 8.95 + 26.51 – 12.72). This interaction term was negative, $\gamma_{02} = -12.72$, $t(57) = -2.42$, $p = .019$, indicating that the differential between the treatment and control groups was larger at Urban Southeast than it was at Suburban South. However, the interaction terms for the other three schools (Urban Magnet, Midwest Suburban, and Southern Rural) were not statistically significantly different from 0.

The predicted posttest reading comprehension score for a fifth grade male nonminority student of average pretest comprehension score who was exposed to SEM-R treatment at Urban Magnet was 222.45 (198.00 + 8.95 + 25.32 – 9.82). The predicted posttest reading comprehension score for a fifth grade male nonminority student of average pretest comprehension score who was exposed to SEM-R treatment at Midwest Suburban was 217.79 (198.00 + 8.95 – 5.14 + 15.98). The predicted posttest reading comprehension score for a fifth grade male nonminority student of average pretest comprehension score who was exposed to SEM-R treatment at Southern Rural was 213.42 (198.00 + 8.95 – 9.16 + 15.63). Table 13 summarizes the predicted posttest reading comprehension scores for each school.

The coefficient for SEM-R, $\gamma_{05} = 8.95$, $t(57) = 2.24$, $p = .029$, represents the difference mean between treatment group and the control group on posttest reading comprehension for a Urban Southeast student whose pretest comprehension score is at the grand mean, after controlling for gender, ethnicity, pretest reading comprehension score, and grade level. Thus, the predicted posttest comprehension score in the treatment group at Urban Southeast for a male nonminority fifth grade student of an average pretest

reading comprehension score was 206.95 (198.00 + 8.95), whereas the predicted score for a similar student in the control group was 198.00.

To determine which schools had statistically significant treatment effects, we reran the analyses using each of the five schools as the reference school. The results suggested that although there was a positive treatment effect for Urban Southeast, there were no differences between the treatment and control groups at any of the other four schools. Table 10 shows the effect sizes for the difference between the treatment and control groups for each of the schools in Cohen's *d* units. The standardized mean differences between the treatment and control groups ranged from low of –.11 at Suburban South to a high of .27 at Urban Southeast.

Finally, we examined the correlation between reading fluency and reading comprehension. The correlation between our fluency measure and the standardized ITBS score was .80 during the preassessment and .73 during the postassessment.

In summary, we found statistically significant differences favoring the SEM-R treatment group in ORF and in reading comprehension at the high-poverty urban school (Urban Southeast). SEM-R students outperformed control students in reading fluency at Suburban South. However, there were no appreciable differences between the treatment and control groups at the other three schools. We did however find that we could eliminate 4 to 5 hours of regular whole class or group instruction and replace this time with SEM-R activities, including differentiated conferences at all schools with no decrease in fluency or comprehension scores in four schools and increases in the high-poverty urban school. These findings do suggest that the enriched reading program with differentiated instruction produced higher ORF in two schools and higher reading comprehension than the standard basal reading program in the urban school in this study, a result that has been consistent across previous research on the SEM-R in urban settings (Reis et al., 2007; Reis et al., 2008).

TRAPS

The analyses of the TRAPS identified that teacher attitudes regarding reading were similar in the two groups (treatment and control). On the attitude scale portion of the instrument, on a scale of 1 to 7, within which 7 indicated a more positive attitude toward reading, treatment group teachers had a mean pretest score of 5.64 (SD = 0.43) and control group teachers had a mean pretest score of 5.37 (SD = 0.60). These scores were not significantly different (t = 1.93, df = 57, $p > .05$).

Student Engagement

The most dominant student finding that emerged across observations and teacher interviews related to teacher and observers' findings of

increased student enjoyment and engagement in reading. This theme was consistently observed and discussed in both observations and interviews. Across each site, teachers consistently discussed their belief, in some cases for the first time in years, that the greatest benefit of SEM-R was increased student engagement in enjoyable reading experiences. Across all elementary schools, teachers reported more positive student attitudes toward and greater engagement in reading, and they attributed these differences to their implementation of SEM-R. One of the first changes that teachers reported was the creation of a reading climate that resulted in increased enjoyment of and engagement in reading in their classrooms. Teachers across all schools consistently discussed their perceptions that the use of the SEM-R contributed to the creation of a more enjoyable reading climate for students and cited, during interviews and in their teacher logs, success stories about student enjoyment of and engagement in reading. More than 90% of the teachers and all of the principals interviewed, when asked about the benefits of using SEM-R, discussed students' enjoyment of and engagement in reading as the most important benefit. Observations across treatment classrooms also demonstrated high levels of student engagement in reading. The second most frequent finding related to teachers enjoyment about their use of a more flexible approach to reading instruction that enabled them to use their professional judgments about differentiated instruction. Teachers described the enjoyment they experienced about not having to teach the same content, in the same way, on a daily basis.

Discussion

This study examined how a reading comprehension program involving differentiated instruction and a focus on engagement in reading influenced children's reading comprehension and fluency, compared to children in regular reading programs in their schools. Results showed that the use of both differentiated instruction and enrichment teaching methods, including high-interest, self-selected books that are above students' current independent reading levels, resulted in higher reading fluency and comprehension in some students. Teachers were able to replace whole and small group instruction with differentiated instruction without detriment to achievement scores. Although differentiation is widely acknowledged to be an important instructional approach for all children (Tomlinson, 2001), little experimental research has examined the use of differentiated instruction in reading. In the SEM-R, teachers eliminated whole group, primarily basal reading instruction, for an hour each day, and replaced it with individually selected independent reading selections accompanied by 5-minute differentiated teacher conferences, which resulted in the same or higher reading fluency and comprehension scores for the students in the SEM-R group as compared to the control groups. Although differentiation is widely recommended (Tomlinson, 2001;

Renzulli & Reis, 1997), no other experimental research was found that explored the use of differentiation as opposed to whole or small group reading instruction in five diverse elementary schools. The use of the SEM-R in the high-poverty urban school in this study resulted in reading fluency and comprehension scores that were statistically significantly higher from those of the control group, a result that has been noted in smaller scale studies using the SEM-R in urban schools (Reis et al., 2007; Reis et al., 2008; Reis & Housand, 2009). It is also important to note that none of the control classrooms in any of the schools significantly outperformed the SEM-R classrooms, suggesting that differentiated instruction and content works as well or better than regular curricular instruction and content. These combined results suggest that potentially up to 4 to 5 hours of weekly grouped reading instruction could be successfully replaced by brief, differentiated, targeted reading instruction delivered in weekly independent reading conferences with individual students conducted during 35 to 45 minutes of daily independent self-selected reading of content that is above students' current level of reading. These results suggest that the SEM-R approach works as well as or better than the more traditional reading instruction used in these schools, reflecting the usefulness of theoretical principles regarding differentiated learning and individualized support. Moreover, this research suggests that the pattern of similar effects on achievement held for different groups of students within the total sample; although girls outperformed boys and White and Asian students outperformed their counterparts in other minority groups, these patterns were similar in the treatment and control groups. In the future, a longer intervention will be implemented, as will additional professional development. Also, we hope to investigate the use of SEM-R with additional urban and high-poverty schools.

The reading fluency and comprehension results for high-poverty students in Urban Southeast School were significantly higher than their control group peers. Each of the components of the SEM-R framework focuses on engagement, and the SEM-R students at Urban Southeast may have outperformed their control group peers because of the increased engagement and interest they experienced when given an opportunity to choose to read books of appropriate challenge in areas of personal interest (Greenberg et al., 2006; Guthrie & Alao, 1997; Ivey & Broaddus, 2001; Renzulli, 2008; Renzulli & Reis, 1997). This type of opportunity may be less available in the homes and schools of these high-poverty urban students. The increased reading achievement at this school may also have been because of the SEM-R focus on engagement (Greenberg et al., 2006; Guthrie & Alao, 1997; Ivey & Broaddus, 2001; Renzulli, 2008; Renzulli & Reis, 1997) and differentiation as well as scaffolding of advanced thinking skills and higher order questioning as opposed to the direct instruction, test preparation, and remedial focus of the control group instruction (Clark & Graves, 2005; Hobsbaum, Peters, & Sylva, 1996; Rogoff & Wertsch, 1984; Wood & Wood, 1996). In this urban

school, the control classrooms were largely dominated by skills practice, test preparation instruction, and whole group reading instruction with minimal time spent reading. In addition, the environment created by the SEM-R approach to reading may have enhanced children's desire to read and maintain their engagement in reading, as compared to the environment in control classes, as suggested by the qualitative findings discussed (Garan & DeVoogd, 2008).

This study may also raise some financial considerations about the costs of some of the basal programs and remedial instructional programs used in the control classrooms. Estimates of the costs for SEM-R are approximately $1,000.00 per classroom library because of the discounts on the sets of books provided to the SEM-R treatment teachers. This price carries over to subsequent years as teachers can continue to use this classroom library, so expenses will decrease over time. Professional development costs were limited to a day of in-service and the assistance of literacy coaches within each school that principals considered part of their professional responsibilities. This means that the per pupil price for the first year for the books ranges from about $40 to $45 per student for the first year. According to comparison cost figures compiled by Schacter (1999) in an independent evaluation of reading programs that work, the expenses for SEM-R would be less than half the price of DISTAR, Open Court, and Great Books, one quarter of the price of programs such as Success for All and of some of the basal programs used in the schools that participated in this study, if one adds the cost of the ancillary book selections and workbooks that were expended by the participating schools. Some of the administrators, in exit interviews, commented on the cost savings they would experience if they were to enhance the classroom libraries for SEM-R and used their older basal programs to augment instruction in the other half of the double block of language arts instruction as opposed to ordering new basal programs.

Limitations

This study has limitations to be considered. With the use of random assignment within a school setting, there is always a possibility of treatment diffusion from treatment to control classes. Research team members monitored control classrooms during observations for treatment diffusion, and none was noted. Issues of treatment fidelity are also potential limitations for any research study. Treatment fidelity was addressed by conducting unannounced regular observations in each treatment classroom. Detailed notes regarding treatment fidelity were created during each treatment classroom observation, but the use of differentiation and individual reading strategies makes the degree of fidelity more difficult to quantify. Another limitation for this research study results from the design of the three phases of the SEM-R intervention. Three different types of learning activities are

used in this approach, and differences in student achievement cannot be attributed to any one activity. There are other potential reasons for the lack of statistical significance in the other schools in this study including measurement issues, but these are the realities in experimental research in different schools. The absence of a measurement for student engagement is a limitation as data on engagement were collected from teacher interviews and questionnaires in exit interviews. Finally, the lack of data collected on SES at the student level is a limitation of the study. While all schools qualified as Title 1, there was variation within schools in student SES level based on the free and reduced lunch percentages at each school.

Future Research

This study demonstrated that the use of an enrichment reading approach that resulted in high student engagement, coupled with differentiated instruction and a resulting reduction of whole group instruction, was as effective as or more effective than a more traditional whole group basal approach to reading instruction. These conclusions, as well as the limitations of this study, suggest interesting possibilities for future research. In future SEM-R research, we hope to investigate the use of this approach for a full academic year, as opposed to 5 months, and also to explore whether differential benefits occur for urban students and students of poverty. A measure of student engagement in reading will also be included in future research. Additional future research includes working with urban middle schools, a close examination of student logs, a focus on subgroups such as identified gifted students and students receiving special education services, and investigating book choices and the use of reading strategies.

Note

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number S206A040094, as administered by the Office of Elementary and Secondary Education, U.S. Department of Education. The findings and opinions expressed in this report do not reflect the position or policies of the Office of Elementary and Secondary Education or the U.S. Department of Education.

References

Allington, R. L. (2001). *What really matters for struggling readers: Designing research-based programs*. New York, NY: Addison-Wesley Longman.

Allington, R. L. (Ed.). (2002). *Big brother and the national reading curriculum: How ideology trumped evidence*. Portsmouth, NH: Heinemann.

Allington, R. L., & McGill-Franzen, A. (1989). School response to reading failure: Chapter 1 and special education students in grades 2, 4, and 8. *Elementary School Journal, 89*, 529–542.

Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. (1985). *Becoming a nation of readers*: *The report of the commission on reading*. Washington, DC: National Institute of Education.

Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly*, *23*, 285–303.

Archambault, F. A., Jr., Westberg, K. L., Brown, S. W., Hallmark, B. W., Emmons, C. L., & Zhang, W. (1993). *Regular classroom practices with gifted students: Results of a national survey of classroom teachers* (RM93102). Storrs: University of Connecticut, The National Research Center on the Gifted and Talented.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bryan, G., Fawson, P. C., & Reutzel, D. R. (2003). Sustained silent reading: Exploring the value of literature discussion with three non-engaged readers. *Reading Research and Instruction*, *43*, 47–73.

Byrnes, J. P. (1996). *Cognitive development and learning in instructional contexts*. Boston, MA: Allyn & Bacon.

Chall, J. S., Jacobs, V. A., & Baldwin, L. E. (1990). *The reading crisis: Why poor children fall behind*. Cambridge, MA: Harvard University Press.

Clark, K. F., & Graves, M. F. (2005). Scaffolding students' comprehension of text. *Reading Teacher*, *58*, 570–580.

Cunningham, J. W. (2001). The National Reading Panel report. *Reading Research Quarterly*, *36*, 326–335.

Dowhower, S. L. (1987). Effects of repeated reading on second-grade transitional readers' fluency and comprehension. *Reading Research Quarterly*, *22*, 389–405.

Duffy, G. G., Roehler, L. R., & Herrmann, B. A. (1988). Modeling mental processes helps poor readers become strategic readers. *Reading Teacher*, *41*, 762–767.

Duke, N. (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly*, *35*, 202–224.

Flippo, R. F. (1998). Points of agreement: A display of professional unity in our field. *Reading Teacher*, *52*, 30–40.

Fogarty, E. A., Little, C. A., & Reis, S. M. (2005). *Teaching and reading: Attitudes and practices survey (TRAPS)*. Storrs, CT: Schoolwide Enrichment Model-Reading project, Neag Center for Gifted Education and Talent Development, University of Connecticut.

Fountas, I., & Pinnell, G. S. (2001). *Guiding readers and writers grades 3–6: Teaching comprehension, genre, and content literacy*. Portsmouth, NH: Heinemann.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading comprehension: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*, 239–256.

Gambrell, L. B., Palmer, B. M., Codling, R. M., & Mazzoni, S. (1996). Assessing motivation to read. *Reading Teacher*, *49*, 518–533.

Garan, E. M., & DeVoogd, G. (2008). The benefits of sustained silent reading: Scientific research and common sense converge. *Reading Teacher*, *62*, 336–344.

Greenberg, D., Gilbert, A., & Fredrick, L. (2006). Reading interest and behavior in middle school students in inner-city and rural settings. *Reading Horizons*, *47*, 159–173.

Guthrie, J. T. (2004). Classroom contexts for engaged reading: An overview. In J. T. Guthrie, A. Wigfield, & K. C. Perencevich (Eds.), *Motivating reading comprehension: Concept-oriented reading instruction* (pp. 1–24). Mahwah, NJ: Lawrence Erlbaum.

Guthrie, J. T., & Alao, S. (1997). Designing contexts to increase motivations for reading. *Educational Psychologist*, *32*(2), 95–105.

Guthrie, J. T., Hoa, L., Wigfield, A., Tonks, S., Humenick, N., & Littles, E. (2007). Reading motivation and reading comprehension growth in the later elementary years. *Contemporary Educational Psychology*, *32*, 282–313.

Guthrie, J. T., Hoa, L. W., Wigfield, A., Tonks, S. M., & Perencevich, K. C. (2006). From spark to fire: Can situational reading interest lead to long-term reading motivation? *Reading Research and Instruction*, *45*, 91–117.

Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (*Vol. 3*, pp. 403–422). Mahwah, NJ: Lawrence Erlbaum.

Guthrie, J. T., Wigfield, A., Humenick, N. M., Perencevich, K. C., Taboada, A., & Barbosa, P. (2006). Influences of stimulating tasks on reading motivation and comprehension. *Journal of Educational Research*, *99*, 232–245.

Guthrie, J. T., Wigfield, A., & VonSecker, C. (2000). Effects of integrated instruction on motivation and strategy use in reading. *Journal of Educational Psychology, 92*(2), 331–341.

Hasbrouck, J. E., & Tindal, G. (2005). *Curriculum-based oral reading fluency norms for students in grades 2 through 5* (Fluency Norms Technical Report). Eugene: University of Oregon.

Hertberg-Davis, H. L., & Brighton, C. M. (2006). Support and sabotage: Principals' influence on middle school teachers' responses to differentiation. *Journal of Secondary Gifted Education*, *17*, 90–102.

Hiebert, E. H. (2006). Becoming fluent: Repeated reading with scaffolded texts. In S. J. Samuels & A. E. Farstrup (Eds.), *What research has to say about fluency instructions* (pp. 204–226). Newark, DE: International Reading Association.

Hobsbaum, A., Peters, S., & Sylva, K. (1996, March). Scaffolding in reading recovery. *Oxford Review of Education*, *22*, 17–35.

Hoover, H. D., Dunbar, S. B., Frisbie, D. A., Oberley, K. R., Ordman, V. L., Naylor, R. J., . . . Shannon, G. P. (2003). *The Iowa Tests: Guide to research and development*. Itasca, IL: Riverside.

Ivey, G., & Broaddus, K. (2001). "Just plain reading": A survey of what makes students want to read in middle school classrooms. *Reading Research Quarterly*, *36*, 350–377.

Knapp, M. S., Adelman, N. E., Marder, C., McCollum, H., Needels, M. C., Padilla, C., . . . Zucker, A. A. (1995). *Teaching for meaning in high-poverty classrooms*. New York, NY: Teachers College Press.

Krashen, S. (2002). More smoke and mirrors: A critique of the National Reading Panel report on fluency. In R. L. Allington (Ed.), *Big Brother and the national reading curriculum: How ideology trumped evidence* (pp. 112–124). Portsmouth, NH: Heinemann.

Latz, A. O., Speirs Neumeister, K. L., Adams, C. M., & Pierce, R. L. (2009). Peer coaching to improve classroom differentiation: Perspectives from Project CLUE. *Roeper Review*, *31*, 27–39.

Little, C. A., Fogarty, E. A., & Reis, S. M. (2005). *SEM-R observation scale*. Storrs, CT: Schoolwide Enrichment Model-Reading project, Neag Center for Gifted Education and Talent Development, University of Connecticut.

Manning, G. L., & Manning, M. (1984). What models of recreational reading make a difference? *Reading World*, *23*, 375–380.

Marshall, C., & Rossman, G. B. (1989). *Designing qualitative research*. Newbury Park, CA: Sage.

McAllister, G., & Irvine, J. J. (2002). The role of empathy in teaching culturally diverse students: A qualitative study of teachers' beliefs. *Journal of Teacher Education*, *53*, 433–443.

Meece, J. L., & Miller, S. D. (1999). Changes in elementary school children's achievement goals for reading and writing: Results of a longitudinal and intervention study. *Scientific Studies of Reading*, *3*, 207–229.

MetaMetrics, Inc. (2004). *The Lexile framework for reading, FAQ*. Retrieved from http://www.lexile.com/DesktopDefault.aspx?view=ed&tabindex=6&tabid=18

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Moon, T. R., Brighton, C. M., & Callahan, C. M. (2003). State standardized testing programs: Friend or foe of gifted education? *Roeper Review*, *25*, 49–60.

Morrison, T. G., Jacobs, J. S., & Swinyard, W. (1999). Do teachers who read personally use recommended literacy practices in their classrooms? *Reading Research and Instruction*, *38*, 81–100.

National Center for Education Statistics. (2007). *The nation's report card on reading 2007*. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2007/2007496.pdf

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: U.S. Department of Health and Human Services. Retrieved from http://www.nichd.nih.gov/publications/nrp/smallbook.htm

Olenchak, F. R. (1988). The schoolwide enrichment model in the elementary schools: A study of implementation stages and effects on educational excellence. In J. S. Renzulli (Ed.), *Technical report on research studies relating to the revolving door identification model* (2nd ed., pp. 201–247). Storrs: University of Connecticut, Bureau of Educational Research.

Olenchak, F. R., & Renzulli, J. S. (1989). The effectiveness of the schoolwide enrichment model on selected aspects of elementary school change. *Gifted Child Quarterly*, *33*, 36–46.

Pitcher, S. M., Albright, L. K., DeLaney, C. J., Walker, N. T., Seunarinesingh, K., Mogge, S., . . . Dunston, P. J. (2007). Assessing adolescents' motivation to read. *Journal of Adolescent & Adult Literacy*, *50*, 378–396.

Pressley, M., Wharton-McDonald, R., Mistretta, J., & Echevarria, M. (1998). The nature of literacy instruction in ten grade-4/5 classrooms in upstate New York. *Scientific Studies of Reading*, *2*, 159–191.

Rasinski, T. V. (1990). Effects of repeated reading and listening while reading on reading fluency. *Journal of Educational Research*, *83*, 147–150.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A., Cheong, Y. F., Congdon, R., & Toit, M. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Reis, S. M., Eckert, R. D., Fogarty, E. A., Little, C. A., Housand, A. M., Muller, L. M., . . . Sullivan, E. E. (2009). *Joyful reading: Differentiation and enrichment for successful literacy learning, grades K-8*. San Francisco, CA: Jossey-Bass.

Reis, S. M., Eckert, R. D., McCoach, D. B., Jacobs, J. K., & Coyne, M. (2008). Using enrichment reading practices to increase reading fluency, comprehension, and attitudes. *Journal of Educational Research, 101,* 299–314.

Reis, S. M., Eckert, R. D. Schreiber, F. J., Jacobs, J., Briggs, C., Gubbins, E. J., & Muller, L. M. (2005). *The Schoolwide Enrichment Model Reading Study* (RM05214).

Storrs, CT: The National Research Center on the Gifted and Talented, University of Connecticut.

Reis, S. M. & Fogarty, E. (2006, October). Savoring reading, schoolwide. *Educational Leadership*, *64*(2), 32–36.

Reis, S. M., Gentry, M., & Maxfield, L. R. (1998). The application of gifted education pedagogy to teachers' classroom practices. *Journal for the Education of the Gifted 21*(3), 310-334.

Reis, S. M., Gubbins, E. J., Briggs, C., Schreber, F. R., Richards, S., Jacobs, J., . . . Renzulli., J. S. (2004). Reading instruction for talented readers: Case studies documenting few opportunities for continuous progress. *Gifted Child Quarterly, 48*, 309–338.

Reis, S. M. & Housand, A. M. (2009). The impact of gifted education pedagogy and enriched reading practices on reading achievement for urban students in bilingual and English-speaking classes. *Journal of Urban Education, 6*(1), 72–86.

Reis, S. M., McCoach, D. B., Coyne, M., Schreiber, F. J., Eckert, R. D., & Gubbins, E. J. (2007). Using planned enrichment strategies with direct instruction to improve reading fluency, comprehension, and attitude toward reading: An evidence-based study. *The Elementary School Journal, 108,* 3–24.

Reis, S. M., & Renzulli, J. S. (1989). Developing challenging programs for gifted readers. *The Reading Instruction Journal, 32*, 44–57.

Reis, S. M., & Renzulli, J. S. (2003). Research related to the Schoolwide Enrichment Triad Model. *Gifted Education International, 18*(1), 15–39.

Reis, S. M., Westberg, K. L., Kulikowich, J., Caillard, F., Hébert, T., Plucker, J., . . . Smist, J. M. (1993). *Why not let high ability students start school in January?* (RM 93106). Storrs, CT: The National Research Center on the Gifted and Talented, University of Connecticut.

Renzulli, J. S. (1977). *The enrichment triad model guide for developing defensible programs for the gifted and talented*. Mansfield Center, CT: Creative Learning Press.

Renzulli, J. S. (1988). The multiple menu model for developing differentiated curriculum for the gifted and talented. *Gifted Child Quarterly*, *32*, 298–309.

Renzulli, J. S. (2008). Engagement is the answer. *Education Week*, *27*(43), 30–31.

Renzulli, J. S., Callahan, C. M., Smith, L. H., Renzulli, M. J., & Gay, B. H. (2000). *New directions in creativity*. Mansfield Center, CT: Creative Learning Press.

Renzulli, J. S., & Reis, S. M. (1985). *The Schoolwide Enrichment Model: A comprehensive plan for educational excellence*. Mansfield Center, CT: Creative Learning Press.

Renzulli, J. S., & Reis, S. M. (1994). Research related to the Schoolwide Enrichment Triad Model. *Gifted Child Quarterly, 38*, 7–20.

Renzulli, J. S., & Reis, S. M. (1997). *The Schoolwide Enrichment Model: A guide for developing defensible programs for the gifted and talented*. Mansfield Center, CT: Creative Learning Press.

Reutzel, D. R., Fawson, P. C., & Smith, J. A. (2008). Reconsidering silent sustained reading: An exploratory study of scaffolded silent reading. *Journal of Educational Research*, *102*, 37–50.

Reutzel, D. R., Hollingsworth, P. M., & Eldredge, J. L. (1994). Oral reading instruction: The impact upon student reading development. *Reading Research Quarterly*, *29*, 40–62.

Reutzel, D. R., Jones, C. D., Fawson, P. C., & Smith, J. A. (2008). Scaffolded silent reading: A complement to guided repeated oral reading that works! *Reading Teacher*, *62*, 194–207.

Reutzel, D. R., & Smith, J. A. (2004). Accelerating struggling readers' progress: A comparative analysis of "expert opinion" and research recommendations. *Reading & Writing Quarterly*, *20*, 63–89.

Rogoff, B., & Wertsch, J. (Eds.). (1984). *Children's learning in the zone of proximal development* (New Directions for Child Development No. 23). San Francisco, CA: Jossey-Bass.

Schacter, J. (1999). *Reading programs that work: A review of programs for pre-kindergarten to 4th grade*. Santa Monica, CA: Miken Family Foundation.

Scholastic. (2007). *Leveling resource guide*. Retrieved from http://content.scholastic.com/browse/article.jsp?id=4476

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading failure in young children*. Washington, DC: National Academy Press.

Stahl, S. (2004). What do we know about fluency? Findings of the National Reading Panel. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 187–211). Baltimore, MD: Brookes.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research*. Newbury Park, CA: Sage.

Taylor, B. M., Frye, B. J., & Maruyama, G. M. (1990). Time spent reading and reading growth. *American Educational Research Journal*, *27*, 351–362.

Taylor, B. M., Pearson, P. D., Clark, K., & Walpole, S. (2000). Effective schools and accomplished teachers: Lessons about primary-grade reading instruction in low-income schools. *Elementary School Journal*, *101*, 121–166.

Taylor, B. M., Pearson, P. D., Peterson, D. S., & Rodriguez, M. C. (2003). Reading growth in high-poverty classrooms: The influence of teacher practices that encourage cognitive engagement in literacy learning. *Elementary School Journal*, *104*, 3–28.

Teale, W. H., & Gambrell, L. B. (2007). Raising urban students' literacy achievement by engaging in authentic, challenging work. *Reading Teacher*, *60*, 728–739.

Thompson, S., Ransdell, M., & Rousseau, C. (2005). Effective teachers in urban school settings: Linking teacher disposition and student performance on standardized tests. *Journal of Authentic Learning*, *2*(1), 22–34.

Tivnan, T., & Hemphill, L. (2005). Comparing four literacy reform models in high-poverty schools: Patterns of first-grade achievement. *Elementary School Journal*, *105*, 420–441.

Tomlinson, C. A. (1999). *The differentiated classroom: Responding to the needs of all learners*. Alexandria, VA: Association for Supervision and Curriculum Development.

Tomlinson, C. A. (2001). *How to differentiate instruction in mixed-ability classrooms* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Tomlinson, C. A. (2003). Deciding to teach them all. *Educational Leadership*, *61*(2), 6–11.

Topping, K. J., Samuels, J., & Paul, T. (2007). Does practice make perfect? Independent reading quality, quantity and student achievement. *Learning & Instruction*, *17*, 253–264.

Topping, K. J., Samuels, S. J., & Paul, T. (2008). Independent reading: The relationship of challenge, nonfiction and gender to achievement. *British Educational Research Journal*, *34*, 505–524.

Unrau, N., & Schlackman, J. (2006). Motivation and its relationship with reading achievement in an urban middle school. *Journal of Educational Research*, *100*, 81–101.

VanTassel-Baska, J., & Brown, E. F. (2007). Toward best practice: An analysis of the efficacy of curriculum models in gifted education. *Gifted Child Quarterly*, *51*, 342–358.

VanTassel-Baska, J., & Stambaugh, T. (2005). Challenges and possibilities for serving gifted learners in the regular classroom. *Theory Into Practice*, *44*, 211–217.

Westberg, K. L., Archambault, F. X., Jr., Dobyns, S. M., & Salvin, T. J. (1993). *An observational study of instructional and curricular practices used with gifted and talented students in regular classrooms* (RM93104). Storrs: University of Connecticut, The National Research Center on the Gifted and Talented.

Wigfield, A., & Guthrie, J. T. (1997). *Reading engagement: Motivating readers through integrated instruction*. Newark, DE: International Reading Association.

Wood, D., & Wood, H. (1996). Vygotsky, tutoring, and learning. *Oxford Review of Education*, *22*, 5–16.

Yin, R. K. (2002). *Applications of case study research* (2nd ed.). Thousand Oaks, CA: Sage.

Manuscript received September 9, 2009
Final revision received May 26, 2010
Accepted June 2, 2010